

Форма № ДН-7.02.1

Державний вищий навчальний заклад
«Навчально-науковий інститут комп'ютерних наук і технологій»
Кафедра Прикладної математики та інформатики



РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ
ОНД 1.05 Сучасні методи обробки природно-мовних текстів
(шифр і назва навчальної дисципліни)

Рівень освіти: другий (магістерський)
Спеціальність: 122 Комп'ютерні науки
(шифр і назва спеціальності)
Освітня програма: Комп'ютерні науки
(назва освітньої програми)
Мова навчання: українська

Робоча програма навчальної дисципліни Сучасні методи обробки природно-мовних текстів
(повна назва дисципліни)
для здобувачів вищої освіти за спеціальністю 122 Комп'ютерні науки.

« 19 » вересня 2020 року. – 7 с.

Розробники: ас. каф. ПІМІ Черняк Т.О.

Робоча програма затверджена на засіданні кафедри Прикладної математики та інформатики
(назва кафедри)

Протокол № 11 від. « 1 » лютого 2020р.

Завідувач кафедри ПІМІ д.т.н. проф. Дмитрієва О.А.

« 1 » лютого 2020р.
(підпис) (прізвище та ініціали)

Схвалено науково-методичною комісією галузі знань 12 Інформаційні технології
(шифр, назва)
Протокол № 6 від. « 7 » лютого 2020р.

« 7 » лютого 2020р. Голова (Башков С.О.)
(підпис) (прізвище та ініціали)

1. Загальна інформація

Форма навчання	Денна	Заочна
Статус	Обов'язкова	
Обсяг в кредитах ЄКТС	7	
Обсяг в годинах за навчальним планом, разом: в тому числі:	210	
лекцій:	32	
практичні заняття:	32	
лабораторні заняття:	-	
семінари:	-	
самостійна робота:	146	
Форма підсумкового контролю	Екзамен	
Дисципліну викладають	Асистент кафедри ПМІ Черняк Т.О. https://wiki.donntu.edu.ua/view/Черняк_Тетяна_Олександрівна e-mail: tetiana.cherniak@donntu.edu.ua	

Передумови для вивчення дисципліни: теоретичною базою вивчення навчальної дисципліни є такі дисципліни: «Дискретна математика», «Теорія ймовірностей і математична статистика» та «Чисельні Методи».

2. Мета вивчення навчальної дисципліни

Основними завданнями вивчення навчальної дисципліни «Сучасні методи обробки природно-мовних текстів» – навчити студентів проводити синтаксичний аналіз за допомогою БС-граматики, ТП-граматики, мережевої граматики, штучної нейронної мережі Елмана та RAAM, метода N-грам, проводити морфологічний аналіз за допомогою алгебри скінченних предикатів, системи «Сенс-Текст», проводити морфологічний аналіз за допомогою продукційних правил, семантичної мережі, фреймів, сценаріїв, числення предикатів, штучної нейронної мережі XRAAM.

Як результат вивчення навчальної дисципліни повинні бути сформовані наступні компетентності:

- здатність вчитися і оволодівати сучасними знаннями;
- здатність до пошуку, оброблення та аналізу інформації з різних джерел.
- здатність використовувати системні та базові знання зі штучного інтелекту, що включають в себе штучні нейронні мережі, розпізнавання образів та машинне навчання, експертні системи, нечітка логіка, агентні системи, розуміння природо-мовних текстів, автоматичне реферування;
- здатність застосовувати знання структури, функціонування, технічного та програмного забезпечення інтелектуальних та веб-орієнтованих комп'ютерних систем.

Як результат вивчення навчальної дисципліни повинні бути сформовані наступні програмні результати навчання:

- показати в галузі інформатики й сучасних інформаційних технологій, що включають в себе мови програмування, системи управління базами даних та знань, операційні системи, веб-дизайн;
- продемонструвати вміння виконувати комп'ютерне моделювання процесів та явищ на стадії проектування з використанням сучасних комп'ютерних систем;
- продемонструвати здатність генерувати нові ідеї та уміння обґрунтування нових інноваційних проектів та просування їх на ринку;

– показати здатність зрозумілого і недвозначного донесення власних висновків, знань та пояснень до фахівців і нефаківців, зокрема і в процесі викладацької діяльності. Здатність зрозуміти роботу інших, давати і отримувати чіткі інструкції;

– показати здатність до самостійного вирішення поставлених задач інноваційного характеру (кваліфікаційна робота, курсове проектування), уміння аргументувати і захищати отримані результати та прийняті рішення, зокрема і публічно

– показати уміння обґрунтування та оцінювання інноваційних проектів, знання методик просування їх на ринку, вміння виконувати економетричну та наукометричну оцінки.

3. Очікувані результати навчання

Результати навчання, які базуються на програмних результатах навчання:

- вміння проводити синтаксичний аналіз за допомогою БС-граматики, ТП-граматики, мережевої граматики, штучної нейронної мережі Елмана та RAAM, метода N-грам;
- вміння проводити морфологічний аналіз за допомогою алгебри скінченних предикатів, системи «Сенс-Текст»;
- вміння проводити морфологічний аналіз за допомогою продукційних правил, семантичної мережі, фреймів, сценаріїв, числення предикатів, штучної нейронної мережі XRAAM.

4. Засоби діагностики результатів навчання

Поточний контроль здійснюється під час проведення практичних занять і має на меті перевірку рівня підготовленості студента до виконання конкретної роботи. Форма проведення поточного контролю – усна бесіда за результатами виконання практичних робіт.

Підсумковий контроль проводиться з метою оцінювання результатів навчання та визначається підсумками результатів виконання та захисту практичних робіт по кожному зі змістовних модулів та оцінка розрахункової роботи.

Підсумковий семестровий контроль – екзамен.

5. Критерії оцінювання результатів навчання

Критерії оцінювання мають формулювати порядок оцінювання під час поточного контролю (за результатами практичних, лабораторних, семінарських занять та виконання індивідуальних або групових завдань) та підсумкового контролю.

Пр.1	Пр.2	Пр.3	Пр.4	Інд.завд.	Поточний контроль	Іспит	Максимальний бал
7	7	7	7	12	40	60	100

Примітка: Пр.1, Пр.2 і т.д. практичні роботи;
Сз1, Сз2 і т.д. семінарські заняття;
Лр1, Лр2 і т.д. лабораторні роботи.

Результати підсумкового контролю оцінюються за 100-бальною шкалою та чотирибальною («відмінно», «добре», «задовільно», «незадовільно»).

Відповідність між шкалами встановлюється наступним чином:

Оцінка	
За 100-бальною шкалою	Для екзамену, курсового проекту(роботи), практики, диференційованого заліку, кваліфікаційного екзамену, випускної кваліфікаційної (дипломної) роботи (проекту)
90-100	відмінно
74-89	добре
60-73	задовільно
0-59	незадовільно

6. Програма навчальної дисципліни

6.1. Основні теми дисципліни

Змістовий модуль 1. Введення в граматику мови

Тема 1. Словотвір. Основні поняття словотвору. Частини слова. Способи словотворення. Морфологічні явища в словотворенні.

Тема 2. Словозміна. Частини мови і морфологічні категорії. Морфологічні явища в словозміні.

Тема 3. Синтаксис. Синтаксичні одиниці та синтаксичні зв'язки. Формальна організація пропозиції. Смысловая організація пропозиції. Синтаксичні одиниці та синтаксичні зв'язки. Формальна організація пропозиції. Смысловая організація пропозиції. Базові субстанціональні синтаксеми. Вторинні субстанціональні синтаксеми. Базові предикатні синтаксеми. Вторинні предикатні синтаксеми.

Змістовий модуль 2. Дослідження алгоритмів і методів аналізу мовних конструкцій

Тема 4. Морфологічний аналіз. Декларативний метод. Процедурний метод. Комбінований метод (поєднання декларативного і процедурного). Метод заснований на універсальних морфологічних моделях. Системи рівнянь алгебри скінченних предикатів. Система «Сенс-Текст».

Тема 5. Синтаксичний аналіз. Детерміністський підхід. Граматика безпосередньо складових. Трансформаційна граматика, що породжує. Мережева граматика. Нейромережевий підхід. Процедура навчання (алгоритм зворотного поширення). Стохастичний підхід.

Тема 6. Семантичний аналіз. Продукційні правила. Семантична мережу. Фрейм. Сценарій. Обчислення предикатів. Алгебра кінцевих предикатів. Нейромережевий підхід. Реляційні моделі. Об'єднання морфологічного, синтаксичного і семантичного аналізів.

6.2. Темі практичних (семінарських) занять

№ з/п	Назва теми	Кількість годин	
		Д.ф.н.	З.ф.н.
1	Створення контекстно-залежної граматики Тема 3. Синтаксис Тема 4. Морфологічний аналіз	8	
2	Створення трансформаційної породжуючої граматики Тема 4. Морфологічний аналіз	8	
3	Створення розширеної мережі переходів Тема 5. Синтаксичний аналіз	8	
4	Обчислення ймовірності пропозиції на підставі біграм та триграм Тема 6. Семантичний аналіз	8	
...	Усього годин	32	

6.3. Самостійна робота

№ з/п	Назва теми	Кількість годин	
		Д.ф.н.	З.ф.н.
1	Словотвір	16	
2	Словозміна	20	
3	Синтаксис	20	
4	Морфологічний аналіз	30	
5	Синтаксичний аналіз	30	
6	Семантичний аналіз	30	
	Усього годин	146	

6.4. Індивідуальні та/або групові завдання

Індивідуальні заняття з кредитного модуля складаються: із самостійної роботи над поглибленням теоретичного курсу за матеріалами лекцій, самостійної роботи при підготовці до практичних та лабораторних занять.

Навчальним планом дисципліни визначена розрахункова робота (РР) на тему Синтаксичний аналіз в природномовних текстах, що виконується під час самостійної роботи студента та сприяє більш поглибленому вивченню студентом теоретичного матеріалу, формуванню вмінь використання знань для вирішення відповідних практичних завдань.

Метою розрахункової роботи є формування в процесі досліджень подань про синтаксичний аналіз в природномовних текстах. Ручне трасування генерування мовної конструкції на підставі детерміністського та стохастичного підходів.

7. Література

7.1. Основна

1. Большакова Е.И.; Клышинский Э.С.; Ландэ Д.В.; Носков А.А.; Пескова, О.В.; Ягунова Е.В. Автоматическая обработка текстов на естественном языке и компьютерная лингвистика: учеб. пособие., М.: МИЭМ, 2011. – 272 с.
2. Зубов А.В., Зубова И.И. Основы искусственного интеллекта для лингвистов. – М.: Университетская книга, 2012. – 320 с.
3. Елисеева О. Е. Естественно-языковой интерфейс интеллектуальных систем: учеб. пособие. – Минск: БГУИР, 2010. – 151 с.
4. Bessmertny I. On constructing intellectual systems in ternary logic. Programming and Computer Software, 2014, vol. 40, no. 1, pp. 43–46.
5. Большакова Е.И., Воронцов К.В., Ефремова Н.Э., Клышинский Э.С., Лукашевич Н.В., Сапин А.С. Автоматическая обработка текстов на естественном языке и анализ данных: учебное пособие – М.: Изд-во НИУ ВШЭ, 2017. – 269 с.

7.2. Допоміжна

1. Русская грамматика: [в 2 т.] / под ред. Н.Ю. Шведовой. – М.: Наука, 1982. – Т. 1: Фонетика. Фонология. Ударение. Интонация. Введение в морфемистику. Словообразование. Морфология. – 1982. – 784 с.
2. Современный русский язык / под ред. В.А. Белошапковой. – М.: Азбуковник, 1997. – 928 с.
3. Апресян Ю.Д. Избранные труды: [в 2 т.] / Ю.Д. Апресян. – М.: «Языки русской культуры», 1995. – Т. 1: Лексическая семантика (синонимические средства языка). – М.: «Языки русской культуры», – 1995. – 230 с.
4. Золотова Г.А. Синтаксический словарь. Репертуар элементарных единиц русского синтаксиса / Г.А. Золотова. – М.: 1988. – 300 с.
5. Вихованець І.Р. Граматика української мови. Синтаксис: Підручник / І.Р. Вихованець.. – К., Либідь, 1993. – 368 с.
6. Загнітко А.П. Теоретична граматика української мови: Морфологія: Монографія / А.П. Загнітко. – Донецьк: ДонДУ, 1996. – 437 с.
7. Шабанов-Кушнарченко Ю.П. Теория интеллекта. Технические средства / Ю.П. Шабанов-Кушнарченко. – Х.: Вища шк. Изд-во при Харьк. ун-те, 1986. – 136 с.
8. Мельчук И.А. Опыт лингвистических моделей «смысл-текст» / И.А. Мельчук. – М.: Наука, 1982. – 314 с.
9. Гладкий А.В. Синтаксические структуры естественного языка в автоматизированных системах общения / А.В. Гладкий. – М.: Наука, 1985. – 144 с.
10. Каллан Р. Основные концепции нейронных сетей / Р. Каллан. – М.: Изд. дом «Вильямс», 2001. – 288 с.
11. Евдокимова И.С. Естественно-языковые системы: курс лекций / И.С. Евдокимова. – Улан-Удэ: Изд-во ВСГТУ, 2006. – 92 с.

7.3 Методична

Конспект лекцій з дисципліни «Сучасні методи обробки природно-мовних текстів» для студентів спеціальності 122 Комп'ютерні науки денної та заочної форми навчання / [укл. Є.Є. Федоров, І.В. Ярош, Т.О. Черняк]. – Покровськ: ДВНЗ «ДонНТУ», 2018. – 60 с.

<http://ea.donntu.edu.ua:8080/jspui/handle/123456789/29087>

Методичні вказівки до виконання практичних робіт з дисципліни «Сучасні методи обробки природно-мовних текстів» (у розробці).

Методичні вказівки до виконання розрахункової роботи з дисципліни «Сучасні методи обробки природно-мовних текстів» (у розробці).

8. Інформаційні ресурси

1. Getting started with NLP [Електронний ресурс] – Режим доступу: <https://medium.com/@gon.esbuyo/get-started-with-nlp-part-ii-overview-of-an-nlp-workflow-7baf5948b24> – Заголовок з екрану.

2. Огляд комп'ютерно-лінгвістичних методів обробки природномовних текстів [Електронний ресурс] – Режим доступу: <http://dspace.nbuv.gov.ua/handle/123456789/59840> – Заголовок з екрану.

3. Концептуальні алгоритми виокремлення морфем для реалізації інформаційної технології обробки природномовних текстів [Електронний ресурс] – Режим доступу: <http://www.hups.mil.gov.ua/periodic-app/article/8322> – Заголовок з екрану.

